IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

TITLE: SYNCHRONOUS PLAY-OUT OF MEDIA DATA
PACKETS

INVENTOR: Rüdiger MOSIG

William S. Frommer
Registration No. 25,506
FROMMER LAWRENCE & HAUG LLP
745 Fifth Avenue
New York, New York 10151
Tel. (212) 588-0800

Descripti n

The invention relates to a method, a media source, a media sink and a media processing system to enable a synchronous play-out of media data packets.

A human being uses two parameters of sound to determine the position of the sound source: the amplitude and the phase of the sound. Since the intensity of
5   the sound decreases as it travels through air, the ear further away from the sound source receives a lower sound level than the ear close to the sound source. Further, because sound needs time to travel through air, the ear further away receives the signal later than the closer ear. Experiments have shown that human beings perceive a phase difference between the two channels of larger
10   than 6-20 micro seconds (µs) as a displacement of the sound source and two signals with a phase difference of more than 35-40 milliseconds (ms) are perceived as two distinct sounds.

For audio systems that play-out (emit) audio sound this means that an audio signal belonging to one channel of a multi-channel signal, e.g. a stereo signal,
15   should be played at exactly the same time, i.e. exactly the same moment in time, as all other corresponding audio signals belonging to the same multi-channel signal, e.g. the same stereo signal. In other words, a tight synchronization of different audio output devices, e.g. loudspeakers, is necessary so that the time relation between different channels of a multi-channel signal is met during the
20   output. Similar requirements may also occur in other audio applications like e.g. Dolby Surround Systems or in audio-video applications.

The mentioned tight synchronization must also be fulfilled by digital transmission audio systems, where audio signals are transmitted from the media source to the audio output devices (in the following also more generally referred
25   to as media sinks which include also devices to process a received multi-channel signal in any other way) in form of media data packets (in the following also referred to as media packets). Each audio output device must play-out the sound of a media data packet (play-out the media data packet) at exactly the right time, i.e. at the moment another media output device plays out a
30   corresponding media data packet, e.g. belonging to the same stereo signal, but to another channel. If the media data packets are not played-out well synchronized, i.e. corresponding media data packets of different channels

belonging to the same stereo signal are played-out at different times in different media output devices, the above mentioned problems occur, i.e. the stereo sound is eventually perceived as coming from another direction or eventually even two distinct sounds are perceived (these problems are in the following referred to as

5   hearing distortions).

The Internet Engineering Task Force (IETF) has provided a Transport Protocol for Real-Time Applications (RTP) in its Request for Comments RFC 1889, in the following referred to as RTP. The Real-Time Transport Protocol (RTP) includes a control protocol RTCP which provides synchronisation information from data

10   senders and feedback information from data receivers. Regarding the synchronization of streams for media distribution, this protocol provides so-called Sender Reports (SR) which provide a correlation between a sampling clock and a global clock.

The Sender Reports (SR) are sent from the media source to the media sink(s) and

15   contain two timestamps. One timestamp indicates a moment in time in time units of the local sampling clock (local sampling clock time) and the other indicates the same moment in time in time units of the global clock (global clock time). Both timestamps of the SR are created at the same moment. The assumption is made that the global clock time is available to the media source

20   and the media sink(s) between which the media stream is transmitted. A media sink thus has access to the global clock time and can therefore adjust its sampling clock to the global clock.

The main intention of RTP is to provide means for video conferencing in the Internet and to re-synchronize video and audio that is received in separate

25   streams on the same single media sink. The protocol is not intended to ensure the synchronous play-out of media data packets in separate media sinks of a digital transmission audio system. Therefore, when using this protocol for sending out media data packets to media sinks, the media data packets may not be played-out well synchronized in different media sinks, i.e. media data packets

30   belonging to the same stereo signal may not be played-out at the same moment in different media sinks, e.g. loudspeakers. Thus, the above mentioned hearing distortions may occur when using only RTP for digital transmission audio systems.

The problem of hearing distortions may also result from unreliable and unprecise clock information present in most non real-time source devices like personal computers (PCs) or personal digital assistants (PDAs). These devices assume that the global clock information (global clock time) meets all

5 requirements set by the application scenarios. However, this may not be the case. A non real-time device usually gets an actual time (global clock time) for creating timestamps for media data packets via an external connection, e.g. USB or RS232. Because the bus systems that are generally used for this kind of external connection are not designed to allow a transport with very small

10 guaranteed delivery times, the clock information (global clock time) may loose its accuracy when it is used by the PC or PDA, e.g. to determine a timestamp for a media data packet. This means the global clock time indicated by a timestamp of a media data packet may be wrong with respect to the actual global clock time at which the media data packet is actually sent out. Further, the time difference

15 between two times indicated by two timestamps may vary, even though the time difference between the two corresponding actual global clock times do not vary. The reason for this may be that the time required by the external connection to transport the global clock information to the application may vary. Since the timestamps of the media data packets are generally used by the media sinks to

20 determine a play-out time for each packet, the inaccurate and statistically varying time indicated by the timestamps of the media data packets may lead to the mentioned hearing distortions, since media data packets belonging to the same stereo signal may be played-out at different times by the different media sinks.

25 It is an object underlying the invention to provide a media source, a media sink, and a media processing system to enable the synchronous play-out of media data packets, as well as corresponding methods according to which these devices work, so that hearing distortions are avoided when sound is played by different media sinks.

30 A media source to solve the object of the invention according to a first embodiment of the present invention is defined in claim 1, a media sink is defined in claim 5, and a media processing system is defined in claim 10. Further, corresponding methods according to the first embodiment are defined in claims 17, 21, and 26. Preferred embodiments thereof are respectively defined

35 in the respective following subclaims. A media source to solve the object of the

invention according to a second embodiment of the present invention is defined in claim 11, a media sink is defined in claim 14, and a media processing system is defined in claim 16. Further, corresponding methods according to the second embodiment are defined in claims 27, 30, and 32. Preferred embodiments

5   thereof are respectively defined in the respective subclaims.

Therefore, the object of the invention is solved by two different embodiments, having a common inventive idea for the solution. In both embodiments a common play-out time is determined and associated to each media data packet and the media data packet is played-out by a media sink exactly at this common

10   play-out time. In the first embodiment, the common play-out time is determined by the media sinks by adding a play-out time offset to the time indicated by a timestamp of a media data packet. The play-out time offset is determined by the media source and transmitted to the media sinks. In the second embodiment, the common play-out time is determined by the media source for each packet

15   and sent out together with each media data packet in form of a corresponding timestamp.

Solution according to the first embodiment of the invention:

A media source according to the invention is capable of sending out time-stamped media data packets, in particular to one or more receiving media

20   sink(s) as defined below, the timestamp of each media data packet being indicative for the time of creation of the respective media data packet, adapted for determining a play-out time offset, and further adapted for sending out the play-out time offset, in particular to said one or more receiving media sink(s) as defined below.

25   Preferably, the media source according to the invention comprises a sample clock being capable of determining a sample clock time, is capable of determining a global wallclock time, and is adapted for sending out a control packet once in a while, in particular to said one or more receiving media sink(s) as defined below, said control packet comprising two control packet timestamps

30   indicating the same moment in time, the first control packet timestamp of which being measured or defined in time units of said global wallclock time, the second control packet timestamp of which being measured or defined in time units of said sample clock time.

Further, preferably said timestamp of a media data packet is indicative for the time of creation of said time-stamped media data packet in time units of said sample clock time. Also, the media source is preferably designed for sending out
5   the same media data packets to two or more different receiving media sinks.

A media sink according to the invention is adapted for receiving time-stamped media data packets, in particular from a media source as defined above, and further adapted for determining a play-out time offset, precisely determining a
10   global wallclock time, determining a common play-out time for each received time-stamped media data packet by adding the time indicated by the timestamp of said timestamped media data packet and said play-out time offset, and playing-out each received time-stamped media data packet exactly when the determined common play-out time for the received time-stamped media data
15   packet is reached.

Preferably, the media sink is adapted for receiving said play-out time offset once, in particular from a media source as defined above, and for negotiating said play-out time offset with at least one other media sink. Alternatively, the media
20   sink is capable of receiving a control packet, in particular from a media source as defined above, containing a first contol packet timestamp indicating a certain moment in time measured or defined in time units of a sample clock time and a second control packet timestamp indicating the same certain moment in time measured or defined in time units of a global wallclock time, and of converting a
25   time indicated by a timestamp of a time-stamped media data packet measured or defined in units of a sample clock time into a time measured or defined in units of a global wallclock time, based on the information of the first and second control packet timestamp.

30   In a preferred embodiment, the media sink comprises a buffer which is adapted for storing media data packets until said common play-out time is reached.

A media processing system according to the invention comprises a media source as defined above and a media sink as defined above.
35

A method according to the first embodiment of the invention, intended for a media source, comprises the steps of sending out time-stamped media data

packets, in particular to one or more receiving media sink(s), the timestamp of each media data packet being indicative for the time of creation of the respective media data packet, determining a play-out time offset, and sending out the play-out time offset, in particular to said one or more receiving media sink(s).

5   Preferably, the following steps are performed: determining a sample clock time, determining a global wallclock time, and sending out a control packet once in a while, in particular to said one or more receiving media sink(s), said control packet comprising two control packet timestamps indicating the same moment in time, the first control packet timestamp of which being measured or defined

10   in time units of said global wallclock time, the second control packet timestamp of which being measured or defined in time units of said sample clock time.

It is further advantageously, that said timestamp of a media data packet is indicative for the time of creation of said time-stamped media data packet in

15   time units of said sample clock time. Further, it is preferable, that the same media data packets are sent out to two or more different receiving media sinks.

A method according to the first embodiment of the invention to enable the synchronous play-out of media data packets, intended for a media sink,

20   comprises the following steps: receiving time-stamped media data packets, in particular from a media source, determining a play-out time offset, precisely determining a global wallclock time, determining a common play-out time for each received time-stamped media data packet by adding the time indicated by the timestamp of said timestamped media data packet and said play-out time

25   offset, and playing-out each received time-stamped media data packet exactly when the determined common play-out time for the received time-stamped media data packet is reached.

Advantageously, said play-out time offset is received once, in particular from a

30   media source or it is negotiated with at least one other media sink.

Further, advantageously, the following steps are performed, receiving a control packet, in particular from a media source according to anyone of claims 1 to 4, containing a first control packet timestamp indicating a certain moment in time

35   measured or defined in time units of a sample clock time and a second control packet timestamp indicating the same certain moment in time measured or

defined in time units of a global wallclock time, and of converting a time indicated by a timestamp of a time-stamped media data packet measured or defined in units of a sample clock time into a time measured or defined in units of a global wallclock time, based on the information of the first and second

5   control packet timestamp.

Advantageously, the media data packets are stored in a buffer until said common play-out time is reached.

10   Within a method according to the first embodiment of the invention to enable the synchronous play-out of media data packets, intended for a media processing system, the steps of the method intended for a media source as defined above and the steps of the method intended for a media sink as defined above are performed.

15   Solution according to the second embodiment of the invention:

A media source to solve the object of the invention according to the second embodiment of the invention is adapted for determining a play-out time offset and for determining a common play-out time by adding the determined play-out time offset to a current time, and is adapted for sending out time-stamped media

20   data packets, in particular to one or more receiving media sink(s) as defined below, the timestamp of a time-stamped media data packet being indicative for said common play-out time of the media data packet.

Preferably, the media source comprises a sample clock being capable of determining a sample clock time, and is adapted for calculating said current

25   time by reading a global wallclock time only once and adding time periods given by said sample clock to the only once read global wallclock time. Further, preferably, the media source is adapted for sending out the same media data packets to two or more different receiving media sinks.

30   A media sink according to the second embodiment of the invention is adapted for receiving time-stamped media data packets, in particular from a media source as defined above, is capable of precisely determining a global wallclock time, and of determining a common play-out time for each received time-stamped media data packet which is the time indicated by the timestamp of the time-stamped media

35   data packet. Preferably, the media sink has a buffer which is adapted for storing

media data packets until said common play-out time is reached.

A media processing system according to the second embodiment of the invention has a media source as defined above for the second embodiment of the invention and a media sink as defined above for the second embodiment of the invention.

A method according to the second embodiment of the invention to enable the synchronous play-out of media data packets, intended for a media source, comprises the following steps: determining a play-out time offset and a common play-out time by adding the determined play-out time offset to a current time, and sending out time-stamped media data packets, in particular to one or more receiving media sink(s), the timestamp of a time-stamped media data packet being indicative for said common play-out time of the media data packet.

Preferably, the following steps are performed, determining a sample clock time, and calculating said current time by reading a global wallclock time only once and adding time periods given by said sample clock to the only once read global wallclock time. Further, advantageously the same media data packets are sent out to two or more different receiving media sinks.

A method to enable the synchronous play-out of media data packets according to the second embodiment of the invention, intended for a media sink, comprises the following steps receiving time-stamped media data packets, in particular from a media source, precisely determining a global wallclock time, and determining a common play-out time for each received time-stamped media data packet which is the time indicated by the timestamp of the time-stamped media data packet.

Preferably, media data packets are stored in a buffer until said common play-out time is reached.

A method to enable the synchronous play-out of media data packets according to the second embodiment of the invention, intended for a media processing system, comprises the steps of the method intended for a media source and the steps of the method intended for a media sink.

Therewith, according to the invention media sinks can play-out media data

packets exactly synchronized, because a common play-out time is determined and associated to a respective media data packet and the media data packets are exactly played-out at this play-out time in each media sink. The exact play-out in time by the media sinks is possible because the media sinks determine

5   the global wallclock time precisely, since they are generally using specific hardware that do not lead to long processing times, i.e. the media sinks are tightly coupled to the global wallclock time. The common play-out time is coupled to a once read global wallclock time, so that there are no time differences between two times indicated by two timestamps of different media

10 data packets as is the case in state of the art systems as mentioned above. The media source according to the invention, on the other hand, might have only limited access to the global wallclock time in terms of accuracy, since the added play-out time offset can be chosen so that this inaccuracy is compensated in any case.

15 The invention and advantageous details thereof will be explained by way of exemplary embodiments thereof in the following with reference to the accompanying drawings in which

Fig. 1   shows an example of a scenario where a media source sends time-stamped media data packets to two media sinks;

20 **Fig. 2**   shows the access of the media source and n media sinks to the same global wallclock time;

Fig. 3   shows a flowchart to illustrate the process of sending media data packets from a media source to two media sinks that receive and process the media data packets;

25 **Fig. 4**   shows an example where a PC is used as media source and two loudspeakers are used as media sinks;

Fig. 5   shows a flowchart illustrating the interaction of the media source and the media sink, wherein control packets according to the RTP standard are used;

Fig. 6   shows a flowchart illustrating the interaction of the media source
              and the media sink according to a first alternative embodiment of
              the invention; and

Fig. 7   shows a second alternative embodiment of the invention, wherein
5   the media sinks negotiate a play-out time offset among themselves.

Figure 1 shows the basic scenario of a media distribution session with two
synchronized media sinks, i.e. a first media sink 1 and a second media sink 2.
The media source 101 transmits a first timestamped media data packet 1021 to
the first media sink 1 and a second timestamped media data packet 1022 to the
10   second media sink 2. The timestamp of a media data packet indicates the time
the media data packet was generated by the source. The first media sink 1 and
the second media sink 2 decode the media data packets in case of encoded data.
The data is then stored in respective buffers, i.e. a first buffer 1041 of the first
media sink 1 and a second buffer 1042 of the second media sink 2 until the
15   common play-out time 105 for the respective packet is reached. This common
play-out time 105 is determined by the media sinks for each packet by adding a
once determined play-out time offset to the time indicated by the timestamp of a
media data packet. If the common play-out time 105 for a packet is reached the
media data packet is played-out by the media sink. In the example of Fig. 1, the
20   timestamps of the first media data packet 1021 and the second media data
packet 1022 indicate the same moment in time. Therefore, these media data
packets are played-out out by the first media sink 1 and the second media sink
2 at exactly the same moment.

The play-out time offset has to be negotiated among the media source 101 and
25   all sinks of a media session (here the first media sink 1 and the second media
sink 2), taking into account the transmission time periods, the decoding time
periods, the available buffer sizes, and an eventual lax synchronisation of the
media source 101 to a global wallclock time.

For clocks, it is assumed that in a media-streaming device two clocks are
30   available (accessible): the sample clock and the global wallclock. The sample
clock is the clock that is inherent in the media stream. For a CD as an example
of a source of an audio stream this sample clock is running with 44.1 kHz. The
global wallclock can be read by all source and sink devices participating in a

media session. For IP networks, the Network Time Protocol (NTP) describes how a NTP clock can be maintained throughout a network. However, for applications with tight requirements, such as synchronizing two stereo channels, the accuracy and clock resolution of such an NTP clock may not be sufficient.

5   Therefore it is assumed that a clock with much higher accuracy and resolution is available. This is the case in some wireless systems that need a common clock among all peers in order to execute a synchronized frequency hopping. One example for such a wireless system is given according to the Bluetooth specification, where all participants of a piconet maintain a common clock. The

10   time of the common clock can be used by media applications as the global wallclock time. Usually, the sample clock time and the global wallclock time are measured in different units. For example, the global wallclock may tick in units of microseconds, whereas the sample clock may tick in units of single samples as smallest unit.

15   For timestamps, it is assumed that they are principally used the way as described in RTP. This means a timestamp of a media data packet specifies the moment in time the first sample of the packet was created in time units of the sample clock. In addition to the media stream which transmits media data packets according to RTP, control packets are exchanged among the

20   participants, i.e. among the media sources and media sinks of a stream. These control packets contain no media data, but among other information two timestamps indicating the same moment in time, one timestamp indicates the time in time units of the sample clock and the other timestamp indicates the same moment in time in time units of the global wallclock time. With this

25   information, a media sink can determine a sample clock time, if a global wallclock time is given, and vice versa it can determine a global wall glock time, if a sample clock time is given. Therefore, control packets fulfill the function of associating the source sample clock time with the global wallclock time. It is thus possible for a media sink to determine the moment in time a media data

30   packet was generated in time units of the global wallclock time, by converting the time indicated by the timestamp of the media data packet which is given in time units of the sample clock.

In Figure 2 the assumption is made that there is a global wallclock time 201 available to the media source 202 and all n media sinks, i.e. media sinks 203-1,

35   203-2, ... ,203-n. This global wallclock time can for example be the time of the

clock that is used by digital bus systems or wireless digital transmission systems. Furthermore, it is assumed that this clock fulfills the requirements in terms of accuracy and resolution concerning the desired synchronization. Typically, such a clock is located very close to the physical layer, e.g. in the
5   baseband of such a transmission system. For general-purpose devices like a PC or a PDA 202, this means that such a clock is external and can only be accessed via an external connection 204, e.g. USB or RS232. One example for this is a Bluetooth module that is connected to a PC via USB. The Bluetooth baseband clock is synchronized automatically by all devices within a piconet, because this
10   clock information is used to synchronize the frequency hopping of all piconet participants. The native Bluetooth clock information then has to be transported from the Bluetooth module to the PC via the USB bus system.

Because the bus systems that are generally used for this kind of external connection are not always designed to allow a transport with very small
15   guaranteed delivery times, the clock information may loose its accuracy (validity) when it is transported through such a bus system. For example, clock information of an accuracy of a few microseconds looses a lot of its value when it is transported through a bus system that introduces a delay of a few milliseconds, especially when this delay is subject to a random variation of a few
20   milliseconds, i.e. the time that is used e.g. to generate a timestamp may not any longer be valid. In addition, a non real-time operating system 205 that typically is running on general-purpose devices like PCs and PDAs introduces even more uncertainty to the clock information. In Figure 2, the zigzagged arrow 206 through the layers of media source 202 represents this uncertainty and
25   inaccuracy of the clock information that is received by the media application.

On the other hand, as media sinks, i.e. the n media sinks 203-1, 203-2, ... , 203-n, typically single-purpose devices (embedded devices) are used, e.g. loudspeakers. Such a single-purpose device can be implemented as an embedded system. This allows a much more direct path to the global wallclock
30   time, i.e. the global wallclock time can be precisely determined. For example, the media application 207 can run on the baseband processor of the transmission system. This means that the media application can have a very direct access to the clock with no significant delay and no significant uncertainty. Therefore, precise clock information with an accuracy of a few microseconds is available to
35   the media application 207 of the media sink, because it is not transported

through any slow bus system. The straight arrow 208 within the n media sinks 203-1, 203-2, ... , 203-n in Figure 2 indicates this direct access.

As indicated in Figure 2, this invention utilizes the fact that multiple media sinks, i.e. the n media sinks, can be synchronized among themselves very tightly

5 due to their direct access to the global clock, whereas for the source device, a lax synchronization to the sink devices is acceptable. For example, when streaming stereo audio data from a CD-player to two loudspeakers, the delay from sending a packet from the CD-player until it is played-out at the speakers may be a few milliseconds, but the delay between the left and right speaker may

10 only be a few microseconds. Appropriate buffering in the media sinks therefore can compensate the uncertainty of the clock information on the source side. Because the available clock information on the media source is less accurate and less reliable than on the media sink, the synchronization that is achieved on this basis can be called 'Asymmetric Synchronization'.

15 The global wallclock time is preferably only read once on the source side at the very beginning of the streaming session in order to couple the sample clock to the global wallclock time. This clock information can be used to compile the timestamps of the first control packet that is transmitted to the media sinks. For global wallclock timestamps in subsequent control packets, however, the

20 difference in time can then be calculated by counting the number of samples rather than reading the global wallclock time again. This is due to the fact that the variation of the delivery time of the global wallclock information generally is too large and would lead to gaps or jumps in the play-out on the sink side.

In figure 3 the media source 101 sends media data packets to the first media

25 sink 1 and the second media sink 2. At the beginning of the procedure, a play-out time offset has to be negotiated (determined) in a step 304. This negotiated play-out time offset is transmitted to both media sinks, i.e. the first media sink 1 and the second media sink 2, and is further used by the media sinks to determine a common play-out time for each packet. A media data packet

30 timestamp indicates the moment in time a packet was created in time units of the sample clock. To determine a common play-out time, i.e. the moment in time a sink physically has to play-out a media data packet, the time indicated by the timestamp of the media data packet is converted into a global wallclock time in time units of the global wallclock time and the negotiated play-out time offset is

added to this global wallclock time. For the negotiation of the play-out time offset, the expected transmission time, a potential decoding time and the available media sink buffer sizes have to be taken into account. Because the global wallclock time information on the source side can be inaccurate and subject to a statistical variation, the source has to add the worst-case variation time to the play-out time offset. This avoids the situation that the common play-out time has already elapsed once a media data packet reaches the sink.

Even though the global wallclock time of the media source suffers from the above-mentioned variation, it is read once in a step 305 in the beginning of a media streaming session in order to couple the source sample clock to the global wallclock time. In a following step 306 a control packet with two timestamps is transmitted to the first media sink 1 and the second media sink 2. Both of the timestamps of the control packets describe the same moment in time, one timestamp indicates the moment in time in time units of the source sample clock and the other timestamp indicates the moment in time in time units of the global wallclock time. Thus, a media sink, which receives this control packet can determine the moment in time a media data packet was generated in time units of the global wallclock time using the time indicated by the timestamp of the media data packet in time units of the sample clock.

In the next step 307, the media data packets for each sink are compiled and time-stamped with the time of their creation in time units of the source sample clock. In case a separate stream is sent to each sink, this has to be done for each stream. In case that one stream is multicast to multiple sinks, this only has to be done for this one stream. In the example of Fig. 3 there is only one stream that is sent to both media sinks. Therefore, in step 308, a media data packet of the stream is sent out to the first media sink 1 and the second media sink 2.

In the next step 309, each sink decodes the data in case it is encoded data. Also, in this step 309, a sink converts the time indicated by the timestamp of the received media data packet into a time in units of the global wallclock time. Then, each sink determines the common play-out time by adding the negotiated play-out time offset, which is given in units of the global wallclock time, to the converted time indicated by the timestamp of the received media data packet. In the next step 310 each sink buffers the media data until the determined

common play-out time arrives. These buffers in the media sinks have to be large enough to compensate for the random variation of the clock information in the media source, the random variation of the transmission delay and a possible variation of the decoding delay. With the arrival of the common play-out time 105, each sink physically plays out the media in the next step 311.

For any timestamp of subsequently transmitted media data packets from the media source 301 to the first media sink 1 and the second media sink 2, the media source 301 uses the sample clock time in step 312 to determine the time for the timestamp rather than reading the global wallclock time again. This ensures, that no gaps or jumps occur on the sink side due to the inaccuracy of the wallclock time in the source media application. In step 313, the sample clock time is used instead of a read wallclock time for the timestamp for the next media data packet sent from the source to the sink. In step 314 media data packets are sent to each media sink, i.e. the first media sink 1 and the second media sink 2, with the timestamp that indicates the time of their creation given by the source sample clock. The timestamp of a media data packet is generally included in the media data packet as a header information. However, it may also be sent in separate timestamp packets.

As can be seen in figure 3, steps 312, 313, and 314 are repeated until all media data packets of a session are sent out, i.e. new timestamps are calculated, media data packets are compiled with these timestamps and these media data packets sent out to the media sinks.

As a result of such a procedure, the source media application needs only to be coupled loosely to the global wallclock time, whereas each media sink is coupled tightly to the global wallclock time. Therefore, if a non real-time device like a PC or PDA is used as a media source and loudspeakers are used as media sinks, the loudspeakers can be synchronized very tightly among themselves, fulfilling the tight requirements derived from human perception of spatial audio.

Figure 4 shows a possible scenario where the procedure according to the invention can be applied. A Bluetooth equipped PC 400 is multicasting a stereo audio stream in form of media data packets to two Bluetooth loudspeakers, i.e. a first Bluetooth loudspeaker 4021 and a second Bluetooth loudspeaker 4022, via two Bluetooth links, i.e. a first Bluetooth link 4011 and a second Bluetooth link

4012. In each link the media data packets of one audio signal of a stereo signal are transmitted to the respective loudspeaker.

The Bluetooth module on the PC 400 is connected via USB, whereas in the first Bluetooth loudspeaker 4021 and the second Bluetooth loudspeaker 4022
5   Bluetooth is embedded directly into the system design. The global wallclock time to be used by the PC and the Bluetooth loudspeakers is the Bluetooth baseband clock inherent in each Bluetooth baseband implementation. This Bluetooth baseband clock is very well synchronized among all participants of a Bluetooth piconet.

10   The PC 400 as the media source of the audio stream starts with evaluating the quality and delay of the Bluetooth transmission to the first Bluetooth loudspeaker 4021 and the second Bluetooth loudspeaker 4022 using the information that is provided by the control packets as defined in RTP. Further, the PC queries the time needed for decoding and the buffer capabilities from
15   each speaker using appropriate signaling commands. With this information and the random variation of the clock information of the PC, i.e. a maximum possible variation, the PC can determine a play-out time offset. This play-out time offset is transmitted to the first Bluetooth loudspeaker 4021 and the second Bluetooth loudspeaker 4022 once and is added to the time indicated by the timestamp of
20   each media data packet of a media stream to get the common play-out time for each media data packet. In an alternative embodiment of the invetion, a common play-out time may be determined by the media source, here the PC 400, for each media data packet and then transmitted together with each media data packet, as described in connection with Fig. 6 below.

25   The PC 400 as the media source of the stream creates the timestamps. When RTP media data packets are sent, the timestamps in the media data packets describe the moment in time the packet was created in time units of the sample clock. The link to the global wallclock time, here the Bluetooth baseband clock, is achieved by supplying two timestamps for the same moment in time in the
30   RTCP control packets, one timestamp indicating the moment in time in units of the sample clock and the other one in units of the global wallclock, as described above. Because of the inaccuracy of the clock information available on the PC side, however, the baseband clock is preferably actually read only for the first control packet. For consecutive control packets, the time information for the

global wallclock timestamp is created by counting the number of samples passed
since the last control packet and then translating this number of samples into
time in units of the global wallclock. As mentioned above, a control packet has a
global wallclock timestamp indicating a moment in time in time units of the

5   global wallclock time and a sample clock timestamp indicating the same moment
in time in time units of the sample clock time.Therefore, by combining the
information provided by the various timestamps present in the media data
packets and the control packets, each Bluetooth loudspeaker can determine the
moment in time at which a packet was created by the source in time units of the

10   global wallclock time from the timestamp of a media data packet, which
indicates the time of creation in time units of the sample clock. By adding the
negotiated play-out time offset, it is then determined when the samples from
each media data packet have to be played-out. Because each sink can access the
Bluetooth baseband clock directly, all sinks are able to synchronize their sample

15   play-out clocks tightly to the Bluetooth baseband clock.

Because the clock information is imprecise to a certain extent on the source
side, the first Bluetooth loudspeaker 4021 and the second Bluetooth
loudspeaker 4022 as the media sinks of the audio have to compensate for this
inaccuracy with a suitable buffer size. For example, the PC 400 knows that the

20   clock information has a maximum variation of 2 ms. Therefore, in order to avoid
the situation that the play-out time of a media data packet has already elapsed
once the media data packet reaches the sink it includes these 2 ms in the
negotiated play-out time offset. With 2 ms variation, the timestamps created by
the source will be 1 ms too early or 1 ms too late in the worst case. Therefore,

25   the sinks have to provide enough memory to buffer the data for this worst-case
period that is always added by the source device in order to be on the safe side.

Fig. 5 shows a flowchart illustrating the sending process at the media source
101 and the receiving process at the media sink 1 according to the invention,
where the play-out time for each media data packet is determined by the media

30   sink. This example is based on RTP standard. In a first step 603S the play-out
time offset is determined (negotiated) by the media source 101 taking into
account the transmission time periods, the decoding time periods and the
available buffer sizes of the media sinks participating in the media streaming
session. To get these informations, the media source 101 queries the media sink

35   1. The play-out time offset is then in form of a data control packet 604S

transmitted to the media sink 1. This data control packet 604S contains the play-out time offset in time units of the global wallclock. The media sink 1 receives the transmitted data control packet 604R, which corresponds to the data control packet 604S that was sent out by the media source 101. The media

5   sink 1 stores the play-out time offset in units of the global wallclock time in order to determine the play-out time of subsequently received media data packets, as described below.

Before any media data packets are sent from the media source 101 to the media

10   sink 1 a sender report packet is sent from the media source 101 to the media sink 1. Therefore, in the following step 605S, a sender report packet 606S is created. The sender report packet 606S contains two timestamps, a sample clock timestamp 607S indicating a moment in time in time units of the sample clock and a global wallclock timestamp 608S indicating the same moment in time in

15   time units of the global wallclock. The sender report packet 606S is transmitted from the media source 101 to the media sink 1 once in a while. It is transmitted at least once before every media streaming session, however it may also be transmitted in the middle of a media streaming session. The media sink then receives the transmitted sender report packet 606R containing the transmitted

20   sample clock timestamp 607R and the transmittted global wallclock timestamp 608R. Since both of these timestamps indicate the same moment in time, in the following step 609R, the media sink 1 can associate the sample clock time with the global wallclock time. This means, for subsequently received timestamps the media sink 1 can determine a sample clock time, if a global wallclock time is

25   given by the respective timestamp, and vice versa it can determine a global wall glock time, if a sample clock time is given.

In a subsequent data preparation step 609S, the media source 101 creates a media data packet 610S. This media data packet 610S contains a sample clock

30   timestamp indicating the creation time of the media data packet 610S in time units of the sample clock and, further, it contains media data 612S. This media data packet 610S is transmitted to the media sink 1. The media sink 1 receives the transmitted media data packet 610R containg the transmitted sample clock timestamp 611R indicating the creation time of the transmitted media data

35   packet 610R in time units of the sample clock and the transmitted media data 612R. The media sink 1 then, in a first calculation step 613R, calculates the global wallclock time of creation in time units of the global wallclock time using

the information provided by the transmitted sender report packet 606R received earlier. Now, in a second calculation step 614R, the play-out time can be determined in time units of the global wallclock time by adding the calculated global wallclock time of creation and the play-out time offset. Then, in a

5   buffering step 615R, the transmitted media data 612R is buffered until the determined play-out time in time units of the global wallclock time arrives. Finally, in a play-out step 616R, the media data is physically played-out exactly at the determined play-out time, which is now known in time units of the global wallclock time by the media sink 1. As mentioned above, the exact play-out in

10   time is possible, because the media sink 1 has a direct (tight) access to the global wallclock time.

For subsequent media data packets of a media streaming session the data preparation step 609S, the transmission of media data packets from the media

15   source 101 to the media sink 1, the first calculation step 613R, the second calculation step 614R, the buffering step 615R, and the play-out step 616R are repeated. As mentioned above, within such a media streaming session it may also be possible that a sender report packet 606S is sent once in a while from the media source 101 to the media sink 1.

20

Fig. 6 shows a flowchart illustrating the sending process at the media source 501 and the receiving process at the media sink 502 according to a first alternative embodiment of the invention, where the play-out time for each packet is determined by the media source 501 and transmitted with each media data

25   packet. The illustrated process is executed for each media data packet sent from the media source (SRC) 501 to the media sink (SNK) 502. It should be mentioned, that Fig. 6 shows the process at only one media sink 502 participating in a media streaming session. The same process is executed by other media sinks participating in the same media session.

30   In a first step 510 the play-out time is generated for the media data packet 511 that is sent out next. The play-out time depends on the random variation of the clock information of the media source 501, the transmission time periods, the decoding time periods and the available buffer sizes of the media sinks participating in the media streaming session. As above, these informations are

35   negotiated between the media source 501 and the media sink 502. The media data packet 511 contains the media data 513 and the global wallclock

timestamp 512 that indicates the play-out time for the media data packet 511 in units of the global wallclock time. This media data packet 511 is transmitted to the media sink 502. The transmitted media data packet 514 contains the media data 516 and the global wallclock timestamp 515 which correspond to the media

5   data 513, and the global wallclock timestamp 512, respectively that were sent out by the media source 501. After receiving the transmitted media data packet 514, in a processing step 517, the media sink 502 buffers the media data 515 until the play-out time indicated by the transmitted global wallclock timestamp 515 arrives. Then, in a play-out step 518, the media data is physically played-

10   out exactly at the determined play-out time, which is indicated by the global wallclock timestamp 515, by the media sink 502. As mentioned above, the exact play-out in time is possible, because the media sink 502 has a direct (tight) access to the global wallclock time. For user scenarios with tight timing requirements, like synchronizing the left and right channel of a stereo

15   distribution, this access to the clock information is a critical point.

In order to negotiate (schedule) a play-out time, all devices need access to the same clock information (global wallclock time) as a common time reference. Then, the media source can schedule a media data packet and all sinks have to

20   buffer the media data packet until the scheduled global wallclock time has arrived.

Fig. 7 shows a second alternative embodiment of the invention, where the media sinks negotiate the play-out time offset themselves. In this second alternative

25   embodiment, a third media sink 71 and a fourth media sink 72 negotiate a play-out time offset taking into account the transmission time periods, the decoding time periods, the available buffer sizes and an eventual lax synchronisation of the media source 101 to a global wallclock time. The third media sink 71 and the fourth media sink 72 may negotiate the play-out time offset via a direct data

30   link 73 (direct communication channel), or they may negotiate the play-out time offset via a first data link 74 and a second data link 75 over the media source 101 (indirect communication channel). The first data link 74 connects the media source 101 and the third media sink 71 and the second data link 75 connects the media source 101 and the fourth media sink 72. After the play-out time

35   offset is negotiated by the third media sink 71 and the fourth media sink 72, the media source 101 starts sending timestamped media data packets via the first data link 74 and the second data link 75. In the example of Fig. 7 a third

timestamped media data packet 76 is sent via the first data link 74 and a fourth timestamped media data packet 77 is sent via the second data link 75. The timestamps of the media data packets may indicate the time of their creation in time units of the global wall clock time or they may indicate the time of their

5   creation in time units of the sample clock time. In the latter case, a procedure according to Fig. 5 must be executed before media data packets are sent, i.e. a control packet must be sent from the media source 101 to the third media sink 71 and the fourth media sink 72, such that the third media sink 71 and the fourth media sink 72 can determine the time of creation of a media data packet

10   in time units of the global wallclock time. After the third media sink 71 and/or the fourth media sink 72 received a media data packet, they determine the play-out time for the received media data packet by adding the negotiated play-out time offset to the time indicated by the timestamp of the media data packet. In the example, the third media sink 71 determines the play-out time for the third

15   timestamped media data packet 76 by adding the negotiated play-out time offset to the time indicated by the timestamp of this third timestamped media data packet 76, and plays-out the third timestamped media data packet 76 exactly at this determined play-out time. Further, the fourth media sink 72 determines the play-out time for the fourth timestamped media data packet 77 by adding the

20   negotiated play-out time offset to the time indicated by the timestamp of this fourth timestamped media data packet 77, and plays-out the fourth timestamped media data packet 77 exactly at this determined play-out time.

To summarize, according to the present invention, in a digital audio

25   transmission system advantageously media data packets are sent from a media source to media sinks (e.g. loudspeakers). If a media data packet is received by a media sink and contains audio data belonging to an audio signal of e.g. a stereo signal, it is important that this media data packet is played-out at the same moment as a media data packet containing an audio signal of the same stereo

30   signal received by another media sink, i.e. the media data packets must be played-out synchronously. To ensure this synchronous play-out of media data packets in different media sinks, a common play-out time is determined by the media source or the media sink and media data packets are buffered until this common play-out time is reached. The media source or the media sink determine

35   the common play-out time on the basis of a global wallclock time, which is calculated on the basis of a sample clock time.